DEVELOPMENT AND VALIDATION OF A VIDEO-BASED SOCIAL KNOWLEDGE TEST FOR JUNIOR COMMISSIONED ARMY OFFICERS

R. J. Schneider* and J. W. Johnson
Personnel Decisions Research Institutes, Inc.
Minneapolis, MN 55414

ABSTRACT

Social knowledge/skill are increasingly critical to the success of U. S. Army officers. In this paper, we describe development and criterion-related validation of an experimental video-based social knowledge test (SKT) that uses an open-ended response format. This SKT, which measures social knowledge required for junior commissioned officers, overcomes important limitations inherent in other types of social knowledge measures. The limitations overcome by this experimental SKT include: (1) reliance on verbal stimuli that are not truly "social" in nature, and (2) specification of response options from which examinees must choose, thereby limiting ecological validity since effective social behavior usually requires people to generate their own responses. Our SKT was found to have excellent psychometric properties and to correlate at both statistically and practically significant levels with three out of the five dimensions that make up the social performance domain. This SKT, and others that may be developed using similar methodology but different content, shows great promise both for training and selection/classification applications.

1. INTRODUCTION

1.1 Background

Social knowledge/skill are playing an increasingly critical role in the success of U. S. Army officers, the Army's combat readiness, and the Army's ability to carry out its missions. More than ever, junior commissioned officers must possess the attributes necessary to rapidly form and effectively lead small, cohesive units that may have rapidly changing complements of personnel. Officers' insight into their soldiers' anxieties and problems (despite the fact that those soldiers may be reluctant to discuss them), and the soldiers' sense that their leaders are concerned about them are among the critical factors that engender unit cohesion. Moreover, officers must be able to mentor soldiers, work effectively with individuals ranging widely in personality and work style, and be able to adapt to constantly changing mission requirements that may involve deployment to a variety of new cultures. Upon deployment, they may need to establish and maintain relationships not only with diverse groups of soldiers that they have known only a short period of time, but also with indigenous personnel in cultures with value systems and customs very different from their own. Clearly, officers' social knowledge/skill will be instrumental to their effectiveness in these leadership roles.

Given these social knowledge/skill requirements, effective training, selection, and classification based on social knowledge/skill will be essential to the success of the Army's future force leaders. Development of valid psychological tests of social knowledge/skill constructs will, in turn, be instrumental to the development of high-quality training, selection, and classification applications that incorporate social and leadership content. In the present research, we sought to develop such a test. Specifically, we present results of a study in which we developed and validated an experimental video-based social knowledge test (SKT), using an open-ended response format, and based on social episodes derived from a rigorously formulated social performance model.

1.2 Definition of Social Knowledge

To measure social knowledge, one must first define it. We define *social knowledge* as declarative and procedural knowledge/skill necessary for effective social work performance. *Declarative social knowledge* consists of knowledge of people, situations, and social episodes. It consists, for example, of knowledge of the types of behaviors that are appropriate when counseling or helping other military personnel, and the behaviors that typically occur during a performance counseling session. *Procedural social knowledge/skill* consists of rules, skills, and strategies for using declarative social knowledge to construe social events and plan and execute situationally appropriate social action. Successful leaders, for example, use knowledge of how soldiers new to their team are likely to react to various behaviors when they develop strategies to foster unit cohesion (Bartone & Kirkland, 1991).

An important aspect of our definition of social knowledge is its inclusion of the social episode construct. *Social episodes* (e.g., Forgas, 1982) are recurring interpersonal interactions in which a series of goal-directed behaviors unfold over time until (1) the goal is accomplished, (2) something less than full goal attainment is accepted, (3) the goal is determined to be unattainable, or (4) the interactants' attention is directed to one or more other goals (Ford, 1995). Social episodes make an excellent unit of measurement for several reasons. First,

		Form Approved OMB No. 0704-0188

Report Documentation Page

1. REPORT DATE **00 DEC 2004**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Development And Validation Of A Video-Based Social Knowledge Test For Junior Commissioned Army Officers	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Personnel Decisions Research Institutes, Inc. Minneapolis, MN 55414**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001736, Proceedings for the Army Science Conference (24th) Held on 29 November - 2 December 2005 in Orlando, Florida., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**7**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

they integrate knowledge of persons and situations and include a temporal component (i.e., they involve knowledge of persons behaving in situations over time). If one knows a great deal about a social episode one must, therefore, also know a great deal about persons and situations they encompass. Second, job performance is inherently episodic (Motowidlo, Borman, & Schmit, 1997). Therefore, assessing knowledge of social episodes relevant to job performance should provide the best and most efficient prediction of social job performance. Finally, social episodes are "natural" units in the stream of social behavior. As such, it may be easier to capture subject matter experts' (SMEs') expertise regarding social episodes because they are more likely to think in terms of episodes than they are to think in terms of static, de-contextualized persons and situations. Social episodes are closely related to scripts (e.g., Schank & Abelson, 1977), which are cognitive, schema-based knowledge structures that underlie social episodes.

1.3 Approach to Social Knowledge Measurement

In deciding how best to measure social knowledge, we determined that the test should have the following features: First, we wanted the test to be an ability-style measure with right and wrong, or more effective and less effective, answers. Social knowledge is a maximal performance construct, and we wanted to treat it as such. Second, we wanted to experiment with using an open-ended response format, rather than providing people with response options from which to select. We reasoned that, because people usually do not have response options in real-life situations, we might improve on extant measures of social knowledge by not including response options in our test either. Third, we wanted to base our SKT on social episodes because, as described above, job performance is inherently episodic. This construct-matching approach seemed likely to provide more veridical, and therefore more valid, measurement. Finally, we wanted to use video-based social stimuli to enhance realism and minimize spurious overlap with general cognitive ability.

The primary goal of present research was to demonstrate the viability of the social knowledge measurement approach described above by: (1) developing a test according to these guidelines, and (2) evaluating its criterion-related validity against social performance criteria derived from a rigorously developed social performance model. A secondary goal was to formulate such a social performance model and develop a social performance measurement instrument operationalizing that model.

2. METHOD

2.1 Development of Social Performance Model

In order to develop and validate our SKT, we needed to develop a model of social performance requirements for junior commissioned officers. Careful formulation of a social performance model made it possible to specify the content of the SKT, and to develop an instrument to measure social performance constructs serving as the dependent variables in this research. To develop our social performance model, we conducted a literature review that included both scientific and practitioner-oriented literatures relevant to social competence. We also looked at our organization's project files that contained examples of social job performance. This generated over 2,000 social behavior descriptors. We integrated these and selected 291 social behavior descriptors representative of the social performance domain. We then conducted a sorting study, in which 16 psychologists within our organization sorted these 291 descriptors into categories based on their similarity. Results of the sorting study yielded seven social performance dimensions: (1) Teamwork, (2) Coworker Relations, (3) Supervision, (4) Oral Communication, (5) Networking and Customer Relations, (6) Interpersonal Influence, and (7) Interpersonal and Organizational Understanding.

2. 2 Development of Social Knowledge Test

We began the process of developing our SKT by formulating a preliminary list of social episodes adapted from the social behavior descriptors used in the sorting study. We selected 40 from which to extract knowledge requirements. We used the following criteria to select these 40 episodes: We sought to (1) represent the social performance domain specified in our model comprehensively, (2) use social episodes with knowledge requirements that our SMEs would be able to describe accurately, and (3) use social episodes the videotaping of which neither required an excessive amount of money nor imposed undue logistical difficulties.

We held 19 two-hour workshops with a total of 67 3rd- and 4th-year University of Minnesota ROTC cadets and midshipmen[1] to extract these knowledge requirements

---

[1] We used advanced ROTC cadets and midshipmen as SMEs, examinees, and raters in this study. We regarded them as good surrogates for junior commissioned officers, since they are in training to become officers. Moreover, by limiting our study to *advanced* cadets and midshipmen, we ensured that our participants had been socialized into the military to a significant extent, and had been given opportunities to develop and utilize command and leadership skills

from social episodes in our list. For each episode selected for discussion, workshop participants were asked a series of carefully formulated questions:

- What are the main things that usually happen as the social episode unfolds? (e.g., What topics are usually discussed and what actions are usually taken? How do officers usually respond to certain actions?)

- What social norms typically affect officers' behavior during the course of the social episode?

- What are the likely goals and hidden agendas, if any, of the officers in the social episodes?

- What obstacles and challenges commonly arise during the course of each social episode that might hinder an officer's ability to achieve his or her goals?

- What are some effective and ineffective ways of overcoming these obstacles and challenges?

We generated scripts and associated scoring guidelines for 30 of the 40 episodes for which knowledge content was extracted. The 30 episodes were selected based on: (1) the richness and quality of the knowledge content, (2) the likely ease of videotaping the episode, (3) the relative feasibility of writing a script to operationalize the episode, (4) the likely quality and criterion-related validity against important social performance criteria of a social knowledge test item based on the episode, and (5) the need to ensure comprehensive coverage of the social performance domain. Information on which to base the scripts was obtained from: (1) the knowledge extraction workshops described above, (2) various Army and other military websites, and (3) literature relevant to social knowledge requirements for jobs similar to that of junior commissioned officer in the Army.

Scripts were written for the 30 selected episodes. These scripts included not only dialogue, but also "stage directions" to actors to inform them about their characters' motivations and to instruct them to express certain non-verbal behaviors at various points in the episodes. Script paragraphs were numbered to facilitate references to parts of the scripts in the scoring guidelines and discussions of the scripts during various phases of the review and videotaping processes. Finally, a brief scene-setting summary was also written for each script that was included in a voice-over at the beginning of each videotaped episode.

For each episode[2], we wrote scoring guidelines consisting of behaviors targeted as effective and ineffective, reasons for why those behaviors were so targeted, and script paragraph reference numbers that showed where in the scripts the target behaviors were displayed. For each scenario, scores were based on: (1) the number of targeted behaviors identified, (2) the number of reasons identified, and (3) the number of "distracters" identified. In this test, a "distracter" refers to a behavior that might seem ineffective, but really is not; or, conversely, it could be a behavior that might seem effective, but really is not. Points were deducted if an examinee (incorrectly) listed a distracter behavior as either effective or ineffective. This was partly a hedge against examinees who might be inclined to write as many behaviors as possible, hoping that some of them were targeted. It was also another way to evaluate examinees' social knowledge.

Once the scripts and scoring guidelines were created, more workshops were held with ROTC cadets and midshipmen to increase accuracy. We held five such workshops with participants nominated by our ROTC points of contact as high on social and leadership skills. We used a consensus discussion approach, capturing only information on which participants could agree. After the workshops, the commanding officer of an Army ROTC unit conducted a detailed supplemental review of the revised scripts and scoring guidelines, resulting in further revisions.

We pilot tested the SKT on a sample of 22, 3rd- and 4th-year ROTC cadets at the University of Minnesota. Six scenarios were dropped from the pilot test version of the SKT based on pilot test results. Decisions regarding which scenarios to drop were primarily based on review of the number of targeted scoring criteria for a given scenario that differentiated at least somewhat well across examinees. The number of possible points and length of each scenario were also examined to get a sense of the "density of measurement" each scenario contributed to the SKT. "Low-density measurement" in a given scenario meant that the total number of possible points (and, most importantly, the total number of discriminating scoring criteria) per minute was low relative to other scenarios. Item-SKT total correlations also factored into our decisions regarding which scenarios to drop if the item-total correlations of a given scenario with the SKT total-score was substantially lower than that of most other

---

required to perform effectively as junior commissioned officers.

[2] We ultimately decided to refer to these as "scenarios" because we believed that this term would be better understood by examinees and SMEs; for ease of exposition, we will adopt that terminology from this point forward.

scenarios. It was possible that a few of the SKT scoring criteria could have changed as a result of being translated from a written to an audiovisual medium. The first author therefore met with an ROTC cadre officer subsequent to the pilot test to address this possibility and to determine whether certain additional scoring criteria, suggested by the individual scoring the SKT pilot test responses, should be included in subsequent scoring guidelines.

Based on pilot test results and input from the ROTC cadre officer, a 20-scenario SKT was assembled. In this test, examinees view a scenario (which may last anywhere from approximately 1.5 to 4.5 minutes), and then write down all the effective and ineffective behaviors they can identify on an answer sheet. Examinees are given six minutes to do this (the video displays a countdown between each scenario). While identifying behaviors, examinees are provided with the scenario scripts to jog their memory (with stage directions omitted), since social knowledge rather than social memory was the construct of interest. As soon as the six minutes have elapsed, the next scenario appears, and the process is repeated.

2.3 Development of Social Performance Inventory

We developed a multi-source social performance inventory (SPI) to use as our social performance criterion measure in this study. We developed this instrument by adapting the behaviors used in the sorting study. We adapted 71 such statements. We then conducted a small pilot test with three ROTC cadre officers, and reduced the number of items to 52. This was partly due to input from the pilot test that certain items would not work well with an ROTC sample, and partly because of a need to keep the rating process as short as possible, without sacrificing representativeness. We used a 5-point rating scale for this instrument, which assessed the extent to which each behavior characterized a given examinee (with an additional "not observed" option). The SPI also contained written training regarding common rater errors.

2.4 Data Collection

Data were collected on the SKT and SPI from unit personnel in ROTC programs at four U.S. universities. Examinees in this study were limited to advanced cadets/midshipmen (3rd-year and beyond) and junior commissioned officers (Captain and below for Army and Air Force; Lieutenant and below for Navy). In order to obtain a sufficient sample size for our proposed analyses, we collected data from all service branches: Army, Navy Marines, and Air Force. None of the instruments used in this study, including the SKT, was specific to the Army only, so collection of data from all service branches was appropriate. The SKT was administered in group sessions, each lasting 4 hours.

3. RESULTS AND DISCUSSION

3.1 Description of Examinee Sample

The examinee sample was three-quarters male and predominantly white. It was comprised of approximately 50% 3rd-year cadets/midshipmen, 40% 4th-year cadets/midshipmen,[3] and 3% 5th-year cadets/midshipmen. The remaining approximately 6% of examinees were junior commissioned officers. Army was the ROTC service branch with the greatest representation (approximately 44%), Navy and Air Force each constituted approximately 27% of the sample, and about 1% of the sample represented the Marine Corps. Our examinee sample had an average of 1.7 years of prior enlisted military service (SD = 2.6 years). The average age of our examinees was 22.8 years (SD = 2.5 years).

3.2 SKT Scoring

Each examinee received a score for each SKT scenario. This score was (1) the number of effective and ineffective social behaviors, and reasons for their effectiveness or ineffectiveness, that were correctly identified, *minus* (2) the number of distracter behaviors listed. In addition, examinees identifying effective social behaviors as ineffective, or identifying ineffective social behaviors as effective had points deducted. Additional scoring guidelines were articulated in a formal set of *Specific Scoring Guidelines* and *General Scoring Guidelines* that were written based on information acquired from scoring the pilot test examinees' SKTs. The *General Scoring Guidelines* include a description of the SKT, a list of documents to review prior to scoring the SKT, how to use the *Specific Scoring Instructions*, information regarding deduction of points, including the concept of "distracters" as they relate to the SKT, general guidance regarding when to award partial credit, and several other general scoring principles. The *Specific Scoring Guidelines* contain the social behaviors targeted as effective, ineffective, or distracter; reasons why the behaviors are classified as ineffective or effective, and certain additional scoring instructions specific to scenarios and targeted behaviors. Scoring the SKTs was a very labor-intensive process, so the work was split among four Industrial/Organizational (I/O) psychology graduate students. Prior to scoring the SKTs, each scorer was provided with detailed training.

---

[3] This includes two second-year cadets/midshipmen who were exempted from their first two years due to prior military experience

3.3 Inter-Scorer Reliabilities for SKTs

A major concern with regard to the SKT was whether scorers would agree, given the open-ended scoring format. Because of the labor-intensive nature of the SKT scoring process, we investigated the inter-scorer reliability of the SKT by evaluating the extent to which two of four SKT scorers agreed on a subset of 36 examinees. We computed Shrout and Fleiss (1979) Case 2 intraclass correlation coefficients (ICCs) on the profile of 20 SKT scenario total-scores for each of these 36 examinees at both the single-rater and two-rater level. The single-rater ICC is the appropriate reliability measure for those SKTs rated by one scorer only, whereas the two-rater ICC is the appropriate reliability measure for those SKTs rated by two raters. The mean single-rater ICC across the 36 examinees was .83 (SD = .09) and the mean two-rater ICC was .92 (SD = .06). This was considered excellent agreement and indicates that the open-ended scoring approach used for the SKT is capable of producing highly reliable scores when appropriate scorers are used and provided with adequate training.

3.4 Descriptive Statistics for SKT

The scenarios varied considerably in the number of points possible to earn, ranging from a low of 5 points to a high of 24 points (median = 10.5, mean = 11.2, SD = 5.3). Both the mean and median difficulty levels (number of points awarded for a scenario divided by number of points possible) across the 20 SKT scenarios were 0.24. There was, however, good variability in the examinees' scenario total-scores. The mean and median ranges across the 20 SKT scenario total-scores were 0.81 and 0.75 standard deviation units, respectively.

Factor analysis of the SKT scenario total-scores did not yield a coherent structure. We therefore computed a unit-weighted composite of the 20 SKT scenario total-scores ("SKT Composite") so as not to over-weight any aspect of social knowledge. A histogram showing the frequency distribution for the SKT Composite, with normal distribution superimposed, is shown in Figure 1. This figure shows that the distribution of the SKT Composite is approximately normal, and has a range of 4.8 standard deviations.
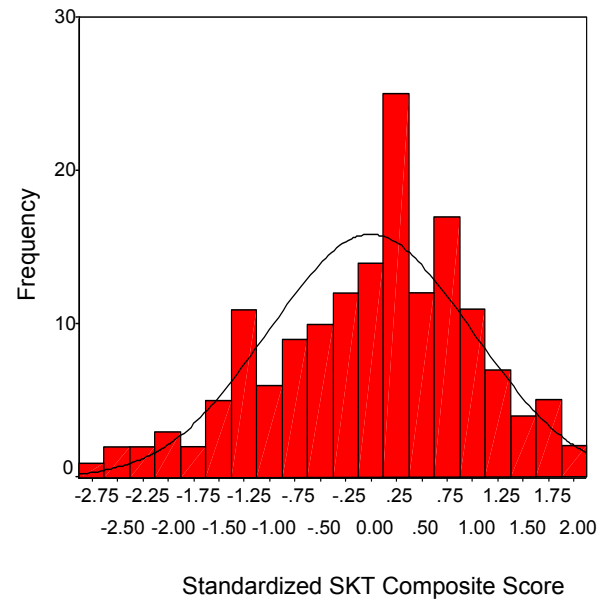

Figure 1. Histogram showing frequency distribution for standardized SKT Composite scores, with normal distribution superimposed.

3.5 Analysis of Social Performance Rating Data

There were 75 raters, who rated a mean of 5.2 examinees each (SD = 2.0), with a range of 1 to 12 examinees per rater. The mean number of raters per examinee was 2.4 (SD = 1.2), with a range of 1 to 5 raters per examinee. Data were aggregated, such that the item scores for each examinee represented the mean rating across raters.

To evaluate the dimensionality of the SPI, we performed a principal axis factor analysis of the SPI items with direct oblimin rotation. A parallel analysis (Horn, 1965) suggested that five factors were appropriate. We assigned the following labels and formulated the following definitions for the factors:

1. *Effective Supervision*: Provides constructive feedback and effectively counsels subordinates; takes into account skills, abilities, and needs of subordinates when working with them

2. *Social Appropriateness*: Does not antagonize, alienate, undermine, betray confidences, or engender feelings of discomfort when interacting with other military personnel; follows military norms regarding appropriate social conduct.

3. *Interpersonal Sensitivity*: Notices when other military personnel are experiencing personal problems/emotional distress, even when their difficulties are expressed obliquely; expresses sympathy and provides support to help them

through these difficulties; develops, maintains, and facilitates good, trust-based working relationships with and among other military personnel.

4. *Handling Social Challenges*: Fits in well when placed in new, interpersonally challenging situations; diffuses tense or uncomfortable social situations with or between others using tactics appropriate to the situation.

5. *Social Presence*: Is persuasive, engaging, and focused around other military personnel; carries self well, with no lapses in military bearing.

The median factor intercorrelation was $r = .25$, indicating that these factors measure distinct aspects of social performance. We computed composites for each SPI factor by computing means of items loading saliently (>.30) on them. With the exception of the Social Appropriateness composite, the SPI factor composites all have means of approximately 3.5 on a 1-5 scale.

In order to estimate the true operational validity of the SKT composite, it is necessary to compute the reliability of the SPI composites serving as dependent variables in this study. We therefore used generalizability theory to estimate the interrater reliability of each SPI composite. Generalizability theory is based on analysis of variance and enables researchers to estimate multiple sources of error variance (e.g., items, raters) within a single design called a generalizability study. The generalizability coefficient, or G-coefficient, represents the ratio of true score variance to true score variance plus all sources of error. The difference between a G-coefficient and a typical reliability coefficient is that many sources of error can be estimated at once, as opposed to only estimating a single source of error at a time (DeShon, 2002). In our study, we had two sources of error variance in the performance ratings: (1) variance due to items, and (2) variance due to raters. Our design was (r: $p) \times i$, or raters nested within examinees and crossed with items. This is because each examinee was rated by a unique set of raters on the same set of items.

To compute the G-coefficient, we conducted an analysis of variance to break the variance in the ratings into the following components: (1) variance due to examinees, (2) variance due to items, (3) variance due to the examinee × item interactions, (4) variance due to the combined rater main effects and examinee × rater interactions, and (5) variance due to undifferentiated rater × item plus examinee × rater × item plus residual effects. We were most interested in the consistency of the relative ranking of persons across conditions, so we computed G-coefficients based on a relative definition of error rather than an absolute definition of error (DeShon, 2002). The relative error term is computed using the following formula (Shavelson & Webb, 1991):

$$\sigma^2_{Rel} = \frac{\sigma^2_{pi}}{n_i} + \frac{\sigma^2_{r,pr}}{n_r} + \frac{\sigma^2_{ri,pri,e}}{n_i n_r}, \qquad (1)$$

where σ^2_{Rel} is relative error variance, σ^2_{pi} is variance due to the examinee × item interaction, $\sigma^2_{r,pr}$ is variance due to the combined rater main effect and examinee × rater interaction, $\sigma^2_{ri,pri,e}$ is variance due to the undifferentiated rater × item plus examinee × rater × item plus residual effect, n_i is number of items, and n_r is number of raters. Because each examinee had a different number of raters, we used the mean number of raters as the value for n_r.

The G-coefficient is computed using the following formula (DeShon, 2002):

$$\rho^2 = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_{Rel}}, \qquad (2)$$

where σ^2_p is variance due to examinee and σ^2_{Rel} is defined as in Equation 1, above. G-coefficients were .40 for Effective Supervision, .52 for Social Appropriateness, .37 for Interpersonal Sensitivity, .51 for Handling Social Challenges, and .63 for Social Presence.

3.6 Evaluating Validity of SKT Composite

Before computing correlations between the SKT composite and the social performance composites, we standardized all of these variables within university. We did this because we found that there were significant mean-score differences between ROTC units from different universities on study variables. This may be because some universities are more selective than others, because some universities have more cadets with prior military experience than others, or because some ROTC units provide more opportunities to acquire social knowledge than others. The problem is that differences between universities on the SKT probably will not translate to similar differences on social performance variables. This is because performance ratings tend to be made on a relative basis rather than an absolute basis. In other words, raters tend to compare the examinee to other cadets/midshipmen with whom they are familiar and make ratings based on how the examinee compares to the norm group. Therefore, the average examinee from one university will likely receive about the same performance rating as the average examinee from another university, even if average performance is much higher at one university than another. To the extent that this happens, the correlation between the SKT composite and the SPI

composites will be attenuated because differences on the SKT composite are not reflected in differences in the performance ratings. By standardizing within universities, mean differences across universities are eliminated, and the correlations between the SKT composite and the SPI composites better reflect their true relationships.

The SKT composite has statistically and practically significant correlations with the Effective Supervision ($r = .30$), Interpersonal Sensitivity ($r = .23$), and Social Presence ($r = .19$) composites (all $p < .01$, one-tailed). When these validity coefficients are corrected for attenuation due to criterion unreliability using the G-coefficients we obtained, validities rise to $r = .47, .38$, and $.24$, respectively. The SKT composite was uncorrelated with the Social Appropriateness and Handling Social Challenges composites. These validities indicate that the SKT shows substantial overlap with social performance dimensions that are critically important for junior commissioned Army officers.

CONCLUSION

The data from this study have provided strong support for the position that a video scenario-based test with open-ended response format is a viable method for measuring social knowledge. We were able to obtain excellent agreement between scorers, and the SKT had good criterion-related validities against three out of five social performance dimensions important to the performance of junior commissioned Army officers: Effective Supervision, Social Presence, and Interpersonal Sensitivity. It was noteworthy that the examinees did not score particularly highly on the SKT, though the frequency distribution showed excellent variability across examinees. Taken as a whole, these data suggest that the SKT would provide an excellent foundation for training applications. It is, of course, possible that junior commissioned officers would have scored more highly on the SKT than ROTC cadets and midshipmen, who are still in training. However, we think it unlikely that junior commissioned officers would score sufficiently highly on the SKT to render it less than useful as a means of both diagnosing training needs and providing a basis for training applications. Moreover, there is no reason that the difficulty level of the test could not be raised or lowered. It also bears mention that this test could be adapted for use with non-commissioned or higher-level officers.

The SKT appears to have considerable promise for diagnosis of critical training needs. In addition, its scoring guidelines, in conjunction with the videotaped scenarios, could be readily adapted into a training module that would facilitate acquisition of social knowledge/skill critical to the success of the Army's future force leaders. We believe that development and evaluation of such training tools would further assist the Army in completing its overall mission and help ensure that its future force will be ready to successfully address the many challenges that undoubtedly lie ahead.

REFERENCES

Bartone, P. T., and F. R. Kirland, 1991: Optimal leadership in small army units. *Handbook of Military Psychology*. R. Gal and A. D. Mangelsdorff Eds., John Wiley & Sons, 393-409.

DeShon, R. P., 2002: Generalizability theory. *Measuring and Analyzing Behavior in Organizations*. F. Drasgow and N. Schmitt, Eds., Jossey-Bass, 189-220.

Ford, M. E., 1995: Intelligence and personality in social behavior. *International Handbook of Personality and Intelligence*. D. H. Saklofske, and M. Zeidner, Eds., Plenum Press, 125-142.

Forgas, J. P., 1982: Episode cognition: Internal representations of interaction routines. *Advances in Experimental Social Psychology*, Vol. 15. L. Berkowitz, Ed., Academic Press, 59-101.

Horn, J. L., 1965: A rationale and test for the number of factors in factor analysis. *Psychometrika*, **32**, 179-185.

Motowidlo, S. J., W. C. Borman, and M. J. Schmit, 1997: A theory of individual differences in task and contextual performance. *Human Performance*, **10**, 71-83.

Schank, R., and R. Abelson, 1977: *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Erlbaum, 248 pp.

Shavelson, R. J., and N. M. Webb, 1991: *Generalizability Theory: A Primer*. Sage, 137 pp.

Shrout, P. E., and J. L. Fleiss, 1979: Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, **86**, 420-428.